Chapter 2

Social Networks and Temporal Social Networks

2.1 Introduction

The role of this chapter is to present the concepts which play the role of underlying layers for dynamic processes studied in Chapter 3. These dynamic processes are diffusion of information, diffusion of innovations and social influence that occur in the network environment. This chapter introduces a number of definitions which serve as a background for these processes. Starting with an idea of event sequence representing contacts or collaborative, social activities between individuals, a definition of a social network is provided. Afterwards, the concept of a temporal social network is defined. It is a time-respecting representation of human interactions in which dynamic processes, such as spread of influence, may be modelled. It makes the whole configuration as close as possible to real-life scenarios.

2.2 Event Sequence

Our typical everyday activities include meeting people, talking to them, exchanging emails, making phone calls, using social media platforms and instant messengers to stay in touch with our friends, relatives and colleagues. Each of those activities often leaves a trace that may be used for analysing multiple aspects of communication. In this thesis, those interactions between individuals are named Event Sequence (ES) and it is defined as follows:

Definition 1. *An Event Sequence is a tuple* $ES = (V^e, EV)$*, where* $V^e = \{v_1^e, \ldots, v_{n^e}^e\}, n^e \in \mathbb{N}_+$ *is the set of social entities and* $EV = \{ev_1, \ldots, ev_{k^{ev}}\}, k^{ev} \in \mathbb{N}_+$ *is the finite set of events (contacts) between them. Each event* ev_{ijkl} *is a tuple:* $ev_{ijkl} = (v_i^e, v_j^e, t_k^e, id_l^e)$*, where* $v_i^e, v_j^e \in V^e$*,* $v_i^e \neq v_j^e$ *and* $t_k^e \in T^e$*. Here,* T^e *represents a discrete time dimension consisting of timestamps* $T^e = \{t_1, \ldots, t_m^t\}$*,* $m^t \in \mathbb{N}_+$ *in which a particular event occurred or is assigned to. The set* $ID = \{id_1, \ldots, id_{n^{id}}\}$ *contains unique event identifiers* $id_{e^{id}} \in ID, n^{id} \in \mathbb{N}_+$*. The set of nodes* V^e *cannot possess any isolated nodes, i.e.* $\underset{i}{\forall}(v_i^e \in V^e \Leftrightarrow \underset{jkl}{\exists}(e_{ijkl} \in EV \vee e_{jikl} \in EV))$*.*

This definition allows to formalize events binding individuals and it introduces the most granular type of information about those events. It means that $ev_{ijkl} = (v_i^e, v_j^e, t_k^e, id_l^e)$ may be interpreted as the fact of sending an email with the identifier id_l^e by an individual v_i^e to v_i^e in time t_k^e. It should be remembered that the content or type of events is not considered here, only the fact of their occurrence. An exemplary ES could be the email server log list or the phone calls registry. Naturally, such an event sequence may be also obtained by conducting surveys and asking respondents to provide the information about who they met and talked to.

In the above definition ID is used for labelling events which were not of unicast type, i.e. when one individual v_i^e contacts towards multiple ones simultaneously, for instance by sending a single email with id_l^e to many recipients. If so, many events ev_{ijkl} are created in ES with the same initiator v_i^e, same timestamp t_k^e and same identifier id_l^e separately for each distinct recipient v_j^e. In order to bind those events and know that they had the same origin, the identifier id_l^e is needed. As sometimes, it is crucial to measure the importance of relations, such a quantitative may be lower if the event was a part of multicast communication and higher if it was a unicast (Kazienko et al. [2009]).

Please note that the set V^e consists only of nodes that have to be a part of any event, being either its initiators or recipients. The restriction $v_i^e \neq v_j^e$ prevents from loops. Events are strictly directed, i.e. if ev_{ijkl} exists, it does not mean that ev_{jikl} also exists.

2.3 Social Network

The event sequence introduced above may be considered as a raw data. By using this as the basis for reasoning, it is hard to perform more sophisticated aggregated analysis. In order to make it more adequate for deeper analysis, it is transformed into the Social Network (SN).

2.3.1 Definition

Definition 2. *A Social Network SN on Event Sequence $ES = (V^e, EV)$ is a tuple $SN = (V, E)$, where $V = \{v_1, \ldots, v_n\}, n \in \mathbb{N}_+$ is the set of vertices and $E = \{e_1, \ldots, e_{k^e}\}, k^e \in \mathbb{N}_+$ is the set of edges between them. Each vertex $v_i \in V$ represents an individual v_i^e from Event Sequence and each edge e_{ij} corresponds to the directed social relationship from v_i to v_j, such that $E = \{(v_i, v_j, w_{ij}) : v_i \in V, v_j \in V, v_i = v_i^e, v_j = v_j^e$ and $\underset{ij}{\forall}(\underset{kl}{\exists} ev_{ijkl} \in EV \Leftrightarrow e_{ij} \in E), w_{ij} \in [0,1]\}$. Here, value $w_{ij} = \frac{n_{ij}^e}{n_i^e}$ denotes the importance (weight, strength) of the relationship between individuals, such that n_{ij}^e is the number of events ev_{ijkl} from v_i^e to v_j^e in ES (regardless k, l) and n_i^e is the number of all events initiated by v_i^e (outgoing from).*

The social network introduced above is defined on the event sequence, i.e. all the nodes that appear in SN have to belong to V^e, actually $V = V^e$, and all the relationships represented by edges E need to be derived from this event sequence. The social network, in fact, aggregates the event sequence file into a directed and weighted graph. Moreover, the weight w_{ij} represents the importance of relationship between v_i and v_j expressed as a fraction of a number of events from v_i to v_j divided by the number of events initiated by v_i.

Due to the fact that the Social Network SN was defined on Event Sequence ES and is based on Definition 1, it is impossible to have isolated nodes in V as well as loops. Moreover, the social network is also directed, i.e. if an edge e_{ij} exists, it does not imply the existence of e_{ji}.

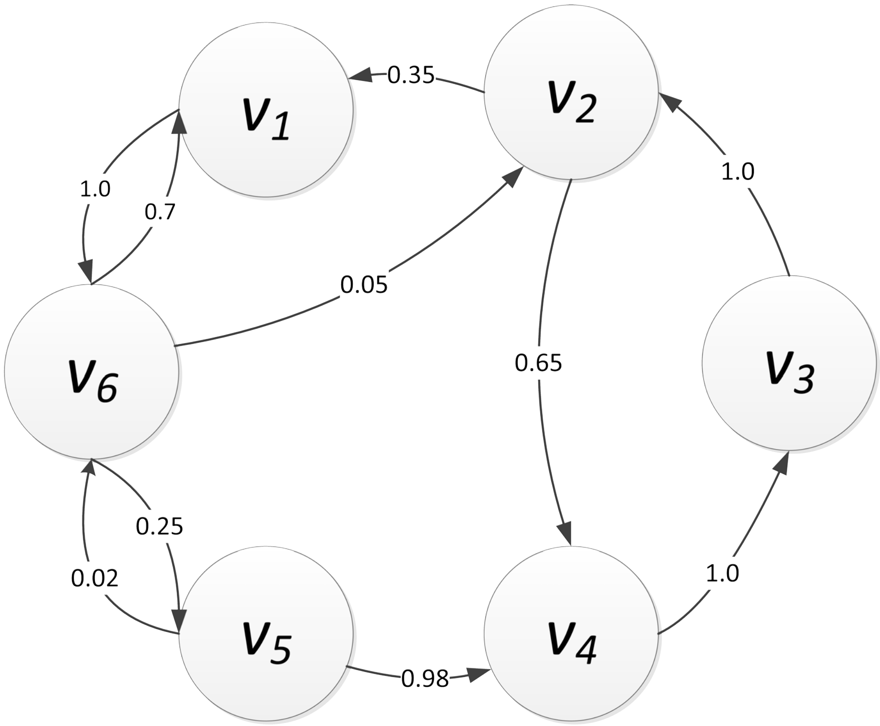

Figure 2.1: An exemplary social network. It consists of six vertices representing individuals and edges reflecting relations. The network is weighted and directed.

This definition enables to perform complex analyses that are mostly derived from the graph theory (Prell [2011]). In the above definition, it is assumed that the relationship is directed, i.e. from v_i to v_j, but there are also variants where the direction is not considered (undirected social networks) or the importance of relation is omitted (unweighted social networks), see Wasserman and Faust [1994]. In this dissertation, further studies are based on the above definition of the weighted and directed social network. An exemplary social network that is compliant with Definition 2 is presented in Figure 2.1.

2.3.2 Limitations

As it may be concluded from Definition 2 of SN, the time dimension is not used here, so it is hard to say whether some edges correspond to old events or recent ones. This is because the timestamp of the event was not considered when transforming the event sequence into the social network. This kind of social network is often called *static* or *time-aggregated* social network (Butts [2009]). In order to overcome this limitation, a number of strategies are proposed. Firstly, instead of aggregating contacts between the same pair of individuals in one edge, multiple timestamped edges are allowed (*contact sequence*, see Holme and Saramäki [2012]). This representation is a straightforward transformation from event sequence to the social network, but it requires the introduction of completely new methods for computing social network measures. The typical ones could not deal with timestamped edges.

Another approach is to create a sequence of static social networks which is time-ordered and every such social network aggregates contacts from a given timeframe (Bródka et al. [2013]). This approach is called a Temporal Social Network (TSN). It could be considered as a trade-off between contact sequence and time-aggregated networks, since it allows the use of standard and well established structural measures while computing the importance of nodes. The temporal aspects are also preserved, because of the time order that is used in a sequence. Before defining TSN, it is required to define a Time-limited Event Sequence (TES) which will simplify the definition of TSN.

2.4 Time-limited Event Sequence

Definition 3. *A Time-limited Event Sequence $TES_{T_p} = (V^e_{T_p}, EV_{T_p})$ on Event Sequence ES is the Event Sequence limited only to events within a given period $T_p = (t_p, t'_p, closure_l, closure_u)$, i.e. $EV_{T_p} = \{(v^e_i, v^e_j, t^e_k, id^e_l) : t^e_k \in T_p\}$. The $closure_l$ refers to the lower bound of the period and can be either of type "[" or "(". Upper bound is described by $closure_u$ and also can be either "]" or ")".*

Please note that this time restriction filters the events to the certain period T_p, but it also refers to nodes, i.e. nodes that are not involved in any activity (event) within the given period (isolated ones) are removed. Obviously, it is possible that if there are no events in ES for the given period T_p, TES will contain empty sets of individuals and events.

Moreover, at this point it is also important to underline one more property of TES. If the period T_p covers all events in ES, the set $V^e_{T_p}$ will be exactly the same as the set V^e. When the T_p is shorter, the set of individuals $V^e_{T_p}$ may not necessarily be the same. This will be the case if the period covered by T_p does not include the individuals that were initiators or recipients only in events outside T_p. In such case $V^e_{T_p} \subset V^e$, but generally $V^e_{T_p} \subseteq V^e$. The same remark applies to EV_{T_p}.

Period $T_p = (t_p, t'_p, closure_l, closure_u)$ may represent four types of ranges: $[t_p; t'_p]$, $[t_p; t'_p)$, $(t_p; t'_p)$, $(t_p; t'_p]$, depending on the closure types. It means that $t^e_k \in [t_p; t'_p]$, $t^e_k \in [t_p; t'_p)$, $t^e_k \in (t_p; t'_p)$, $t^e_k \in (t_p; t'_p]$, respectively.

2.5 Temporal Social Network

Definition 4. *A Temporal Social Network TSN^K on Event Sequence ES is a sequence of time-ordered component Social Networks SN_p, such that $TSN^K = (SN_1, \ldots, SN_p, \ldots, SN_K)$, $K \in \mathbb{N}_+$. A component Social Network SN_p is extracted from Time-limited Event Sequence TES_{T_p} (see Definition 3). The time order is non-descending, i.e.* $\underset{1 \leqslant p < K}{\forall} t_p \leqslant t_{p+1}$ *and* $\underset{1 \leqslant p < K}{\forall} t'_p \leqslant t'_{p+1}$.

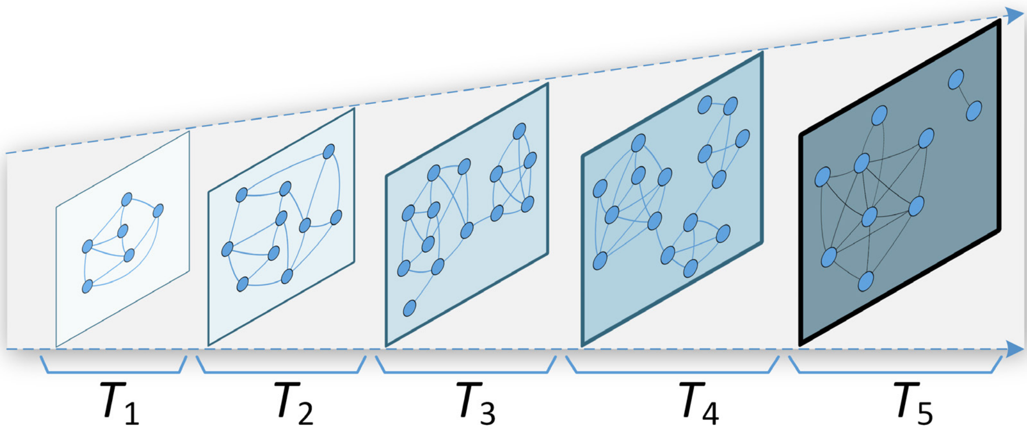

Figure 2.2: An illustration of the temporal social network as defined in this thesis.

Due to the fact that multiple events (common social activities) might occur between v_i and v_j within a single time window T_p, the relationship is obtained by aggregation of these events. Therefore, each SN_p can be treated as a static graph, but the time-ordered sequence of these time-aggregated graphs models the dynamics of the whole network. An illustration of an exemplary temporal social network is presented in Figure 2.2 (this figure was originally introduced in Saganowski et al. [2014]).

Definition 4 does not say that the periods $T_1, ..., T_p, ..., T_K$ have to be of equal length, overlapping or non-overlapping or even whether they have to cover the whole ES. The only condition that has to be satisfied is that $t_1 \leqslant ... \leqslant t_p \leqslant ... \leqslant t_K$ and $t'_1 \leqslant ... \leqslant t'_p \leqslant ... \leqslant t'_K$, i.e. that the beginnings and the ends of the periods have to be in non-descending order. Moreover, as the TES defines two closure types for the beginning of the period and the end of the period, there are four types of periods allowed, see Definition 3.

The most typical variants of TSN are presented in Figure 2.3. Naturally, the choice of how to build a TSN depends on the goal one would like to achieve. As it was presented in Saganowski et al. [2012], the size and type of time windows may influence the results of the analysis of dynamic processes. In this study researchers analysed the group evolution in social networks, but it may also be the similar case for other types of processes. This is why the decision on how to create a TSN based on ES is of great importance. Still, a number of TSN variants seems to be quite natural when thinking about modelling temporal social networks. They are briefly described below.

Non-overlapping consecutive time-windows

The most obvious scenario is presented in Figure 2.3a. Here the time windows are non-overlapping and the whole period of ES is covered by neighbouring social networks following each other. The most important property of this approach is that each event from ES was considered just once, so there is no overlap in particular time windows.

Partially overlapping time-windows

Another way of creating TSNs results with partially overlapping time windows (Figure 2.3b). The reason to create such networks is that there is no strong "cut" that splits social networks. In fact, the border of time windows is a quite a sensitive place where many unique points may be lost, so this is the way to soften this border. One drawback is that here some events will

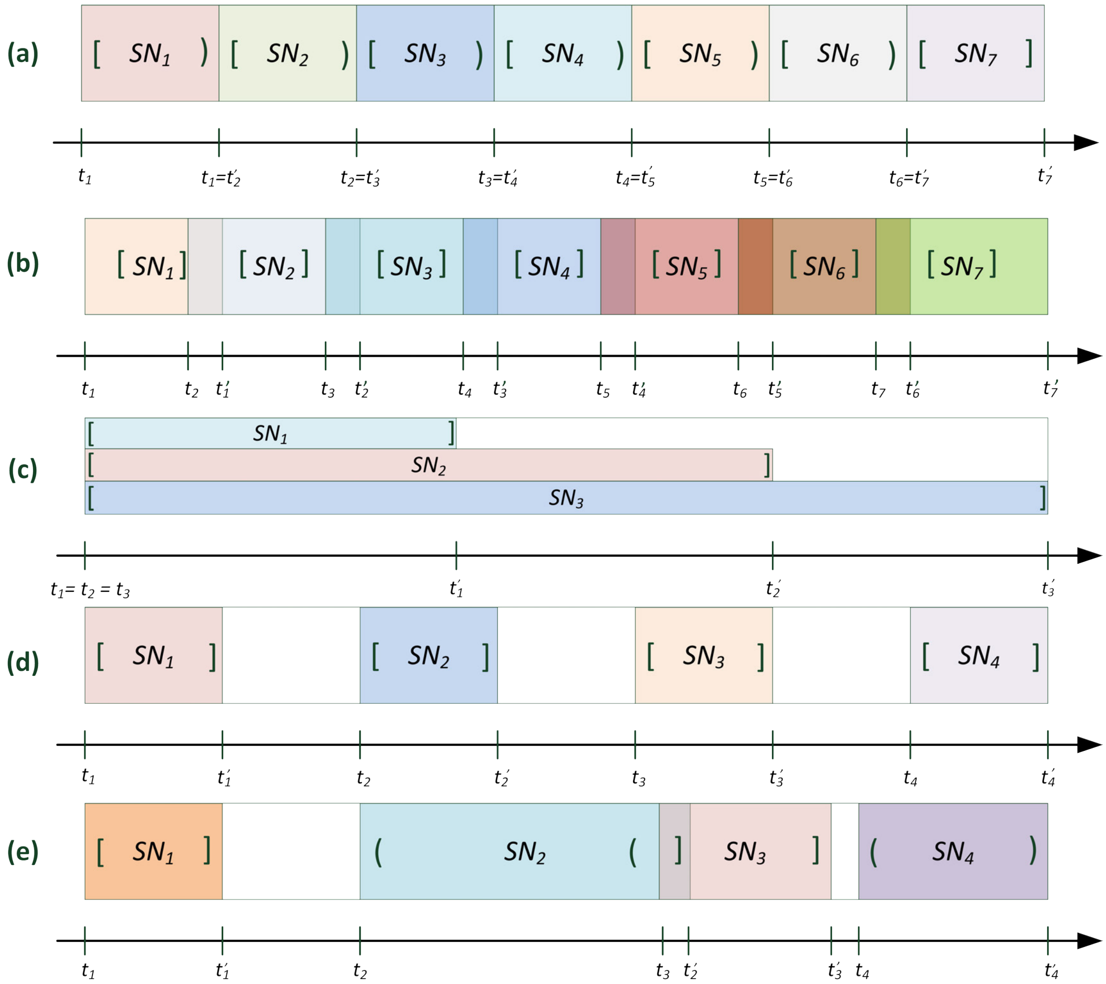

Figure 2.3: Different variants of Temporal Social Networks.

be taken twice, so the edge weights will be also affected.

Increasing time windows

This is a unique approach which, for every social network, uses the events from the beginning of the event sequence, and each social network is built over longer periods (Figure 2.3c). It is used particularly to verify how stable the network is, i.e. whether the structural measures change proportionally to the length of the period. On the other hand, it may be used when the network grows in terms of diameter. This way of building networks may also reveal when some structural holes appear. However, a drawback is most of the burstiness (Barabási [2010]) becomes flattened.

Non-consecutive time windows

In the case of datasets with regular patterns it is often considered whether some periods shouldn't be skipped in order not to disturb this pattern (Figure 2.3d). For instance, if corporate e-mails are analysed and the goal of the analysis is to observe how the network changes on working days, weekends may be omitted. Naturally, a drawback is that some events are not included in the resulting TSN, but in some cases it may be an advantage rather than a loss.

Hybrid approaches

The last approach presented is the mixed one - some periods were longer, some shorter, some overlapping, some not (Figure 2.3e). Basically, the intention of showing this variant here is to show that the process of building the TSN does not necessarily have to be ordered or regular. Sometimes, when trying to capture some special events, it might be reasonable to build a mixed-mode network.

In this dissertation it was decided to build TSNs with time windows of equal size and non-overlapping, where each window is neighbouring with the next one - the case presented in Figure 2.3a. The argument is that each event from ES should be included just in a single SN, as we do not want to lose any event. More details on how the TSNs were created are presented in Chapter 5.

2.6 Discussion

Social networks are created upon human activities. Despite the fact that we are predictable to some extent, the social networks we extract are in fact dynamic. The factors behind the dynamics may be of differing natures, such as meeting new people, changing attitude towards others, changing jobs, moving from one place to another and so on. In fact, the most accurate method of representing human communication is the precise information about who contacted whom at which time - Event Sequence (see Section 2.2). By having this data, it is possible to track how information or influence may potentially flow from one to another. However, without knowing about the content or the essence of the communication it is possible to make just some assumptions about potential paths of information diffusion. Moreover, as a single contact mostly contains just two entities: sender and recipient (directed case, as defined in this dissertation) or contacting parties (undirected case), this granularity cannot benefit from the mature apparatus of Social Network Analysis (SNA). It also requires a lot of storage in comparison to another approach - building a social network from ES, as defined in Section 2.3.1. In such case, the information about communication gets aggregated and the intensity of contacts is most typically expressed as weights over edges (events) between nodes (individuals), as presented in Barrat et al. [2004], Michalski et al. [2011a] or Michalski et al. [2012a]. This approach allows us to obtain a broader view of interconnections in networks and enables us to distinguish groups, hubs, nodes on boundaries of the network and lets us perform other analyses that are offered by SNA techniques (Carrington et al. [2005]; Kazienko et al. [2011]). Unfortunately, as the time-aggregated view of the network is used, all the events are of the same importance, without taking care of their timestamp. From an information or influence propagation point of view, the most important aspect is missing - the order of contacts. As it was stated in Pfitzner et al. [2013], in these networks one assumes transitive paths, which of course do not keep in event sequences or temporal social networks. Moreover, as the contacts within social networks are often *bursty* (Barabási [2010]), the static representation of networks will also ignore this fact leading to wrong conclusions about the dynamic processes taking place in it. This is especially crucial when modelling the spread of epidemics, since the accuracy of predictions may strongly influence the potential actions in healthcare (Masuda and Holme [2013]). To prevent the loss the temporal information, researchers more and more often use temporal representation of networks, and a comprehensive overview of methods of building temporal networks may be found in Holme and Saramäki [2012]. In this work, authors state that the literature on static social networks is many times larger than on temporal ones. This is for a natural reason: it is much easier to analyse time-aggregated networks, especially analytically.

Temporal Social Network, as defined in Section 2.5, is a trade-off between the SN representation and contact sequence representation as defined in Holme and Saramäki [2012]. The author of this thesis agrees with authors of the above cited work that TSN representation may miss some unique points of interaction, but there are also some advantages. Firstly, as the research on temporal social networks is in its early stage, it is hard to name established temporal versions of structural measures. Secondly, the most granular representation also increases computational complexity, since each edge has to be treated individually instead of being aggregated to some extent within the social network - a part of TSN. Lastly, it is far easier to compare the social networks than the contact sequences when trying to answer the question: how do they change in time? Based on this argument, in this thesis temporal social networks are represented as introduced in Definition 4.

2.7 Summary

In this chapter four important concepts that will be used later in this thesis were introduced. Event Sequence ES is an event log which may be treated as a raw data indicating who contacted whom at which time. Then, by using ES a Social Network SN was defined as a time-aggregated static network connecting initiators and recipients of these events, built from nodes and edges. Since the goal was to take advantage of temporal aspects, by defining Time-limited Event Sequence TES (see Definition 3) it was possible to limit the ES to some given period. Finally, Temporal Social Network TSN is a time-ordered sequence of SNs that enables the use of typical SNA techniques in each SN. Temporal aspects are preserved by the appropriate use of TESes. With the introduction of temporal social network, the dynamic aspects of the underlying network layer may be represented. Now, it is time to focus on dynamic processes that take place in those networks.

In the next chapter, a number of those dynamic processes that occur everyday are introduced. Starting with the *diffusion of information*, a process of propagating information through the network, the *diffusion of innovations* is presented. It is a different concept, since it introduces a social change of an individual towards an idea or product. Lastly, *social influence* is described. Social influence is a complex psychological phenomenon which concerns individuals facing a different opinion to their own. The question is whether they will keep their opinion or become influenced by others. Social influence is the process that will be a base for dynamic process taking place in the temporal social network. This dissertation will evaluate whether it is possible to maximize the spread of influence in the temporal environment.

Chapter 3

Diffusion Processes and Influence in Social Networks

3.1 Introduction

As we are immersed in the information era, we cannot avoid contact with information. We receive it from our relatives, newspapers, TV, colleagues and, naturally, from the Internet. The level of exposure to information is as never before, sometimes making us confused. But as the human being is a flexible phenomenon, we start to get used to this information stream. Nevertheless, it is worth analysing how information flows across the social networks and which particular processes are responsible for transmitting it and, sometimes, for changing our mind in some areas.

This chapter aims to present this domain from three different perspectives in order to show the multidimensionality of the problem. Starting from the idea of *information diffusion*, that is the process of information flow in social networks. Second important phenomenon is presented - *diffusion of innovations*. It is the theory of trying to explain why particular ideas have the chance to be adopted while the others do not succeed. This theory studies the process of personal adoption in detail, whereas the last part of this chapter introduces the theory of *social influence* and models of *social influence*. Spread of influence takes a different perspective from the former approach - by using models of social influence at a local network level, the total effect of the process is observed. To conclude the chapter, three approaches are compared synthetically, in an attempt to allow the reader to better understand the similarities and differences between them.

3.2 Diffusion of Information

As it was stated in the abstract of this thesis, people do not live in isolation. It results in the conclusion that by being a part of multiple social networks, we are at the same time the recipients and transmission medium of information. Moreover, apart from our social circles, we may also receive information from external media, such as newspapers, Internet or television, which are sometimes called out-of-network sources (Myers et al. [2012]). This process is symbolically presented in Figure 3.1, where members of a small social network exchange some information (marked as red dotted arrows) over links previously established in the network. Apart from that external sources also spread this information. Each transmission of information is timestamped, since this is a time-respecting process.

Here, the term *information* represents an abstract that may be a simple rumour, it may extend somebody's knowledge, build an opinion on a particular subject or dramatically change someone's life. The process of diffusion of information focuses not on the content of this information, but on the technical capabilities of spreading this information across the social network or, more generally, the society. Trying to observe this process may be extremely challenging task, since it is impossible to have the full knowledge about social network structure, its dynamics and external sources. Moreover, depending on our own perception of the importance of that information, we may initiate contacts to inform our neighbours about it or to pass it when we meet people on different occasion. This distinction is heavily subjective and hard to model. That is why research in this area diverges in a number of directions summarized below. The state-of-the-art in this domain is presented and analysed in Barrat et al. [2008]. In this thesis the concepts of diffusion of information and diffusion of innovations are intentionally distinguished. The reason is that not all information may lead to changing on opinion on something. Nor does all information contain some idea. Some information may simply spread across the network without making any change in its

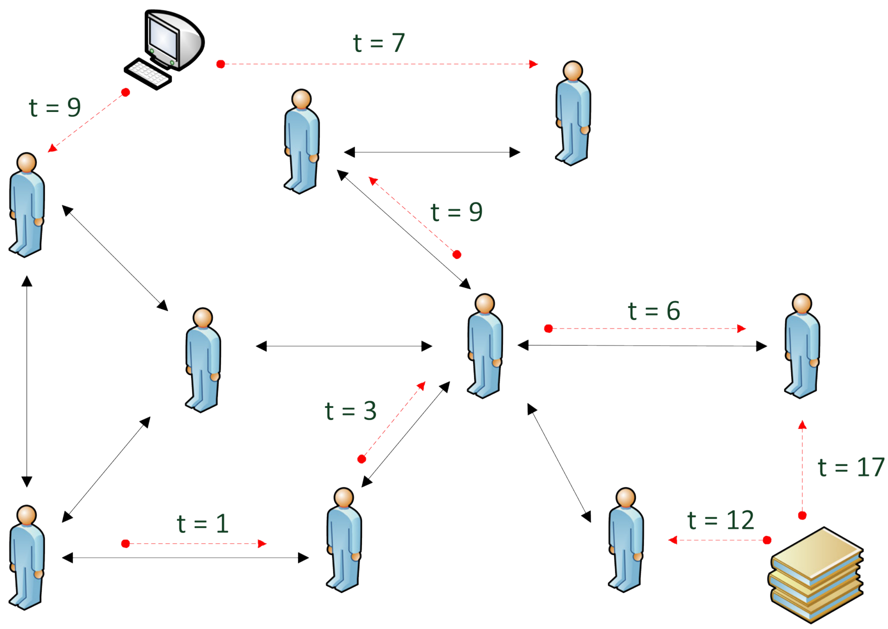

Figure 3.1: The information diffusion occurring in a social network with out-of-network sources (Internet and books). Black arrows represent the underlying network structure and red ones represent the diffusion of information in particular time moments.

state. In contrast, diffusion of innovations is a process where an idea, concept, attitude traverses through the network and changes its members in some way. Nonetheless, some researchers use both terms alternately.

3.2.1 Empirical Research

Researchers attempted to answer the question of whether the networks could be relied upon to transmit information and how long would it take. The most recognizable research is the work by Travers and Milgram [1969] which led to the definition of the small world model by Watts and Strogatz [1998]. This research unveiled that the real world social network has a relatively small average path length, which can be used to predict the speed of information diffusion and its ability to reach different segments of the network. Based on other data, similar evaluations were made of different content and environments, such as: chain-letters (Liben-Nowell and Kleinberg [2008]), blogosphere (Gruhl et al. [2004]), emails (Dodds et al. [2003]) or twitter posts (Yang and Leskovec [2010]). It should be emphasized that a huge acceleration in research in this domain was possible thanks to the development of technology, which allowed the tracking of information flows more precisely. Still, as every social network differs, other results are obtained for each kind of network. Even if the nodes in the network remain the same, the variety of communication methods used to pass the information may also speed up or slow down the diffusion process. To consider this diversity, researchers developed another viewpoint on networks by defining a separate layer of the network for each communication method. This approach is called multi-layered (Jankowski et al. [2013a]; Kazienko et al. [2010]; Magnani and Rossi [2011]) or multiplex networks (Lee et al. [2012]). This kind of network helps in understanding how people are connected by using different mediums or how the edges and communities on different layers overlap. Studying the diffusion in multi-layered networks is more complex, since it requires more data sources to be captured and analysed. For instance, tracking e-mail and mobile communication requires obtaining the data from SMTP servers and telecoms. This is why research in the area is more oriented on simulations rather than on real world experiments (Gomez et al. [2013]; Michalski et al. [2013]).

3.2.2 Models of Diffusion of Information

Another direction in the area of diffusion of information is focused on modelling these processes. An extensive work summarizing the state-of-the art in this area (Bartholomew and Bartholomew [1967]) presents different models for different sociological phenomena and it is a good starting point for discovering the sociological background of particular models. Most of the presented models follow the Markov property (Markov [1951]), some suitable for discrete time and some others for continuous time. The most important from this dissertation's point of view are:

Birth-death processes

Birth-death processes are considered a special case of Markov process, where transactions from one state S_n are permitted only to neighbouring states S_{n-1}, S_n, S_{n+1} (Garcia [1990]). The etymology of the name is straightfor-

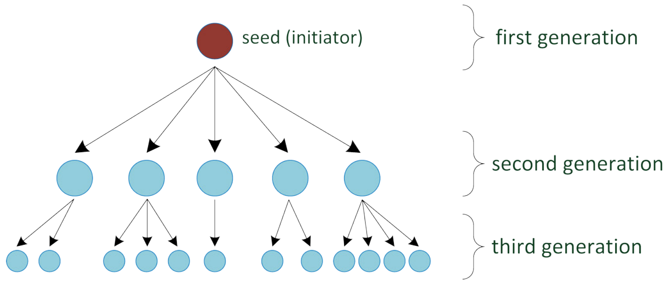

Figure 3.2: The diffusion of viral content represented as a branching process.

ward - these processes may be used to model the population, where people are born and die and the state S_n represents the current population size, λ_n the average birth rate and μ_n the average death rate. In the area of information diffusion these processes may be used for modelling the number of social network members knowing some information but also taking into account the possibility that someone may forget it. Moreover, to consider a process as a birth-death one, it is not realistic to assume that two people may obtain or forget the information at exactly the same time.

Branching processes

Branching processes have Markov properties, but the formulation of the problem is slightly different than in the birth-death problem (Kendall [1966]). Here, it is assumed that the process is parametrized by just one variable μ defining the number of expected children of an individual. It allows the calculation of the expected size of the nth generation, which is μ^n. These processes are widely used to model viral campaigns in social networks (Iribarren and Moro [2011]; Jankowski et al. [2012b]), especially while attempting to discover how "deeply" into the network structure information will spread if the number of neighbours an individual will infect is expressed as μ. An illustration of a viral content diffusion process represented as a branching process is presented in Figure 3.2.

Epidemic models

A novel direction in the modelling of information diffusion has its roots in modelling epidemics. Formal models that are most widely used for predicting disease spread in society are used more and more often to try to model the diffusion of information. The most recognizable ways for modelling the epidemics (see Bailey et al. [1975] as a good reference), such as *Susceptible-Infectious-Susceptible* (SIS) or *Susceptible-Infectious-Recovered-Susceptible* (SIRS) are increasingly often adopted or even directly used in this area. Indeed, some further examination of these models on real world data shows that in some situations they may accurately describe the diffusion of information in social networks (Gruhl et al. [2004]; Xu and Liu [2010]).

There has also been an extensive research conducted in the area of rumour spreading, where researchers try to examine different models that sometimes extend the models presented above, but are more adjusted to model human behaviour (Dietz [1967]; Galam [2003]; Lefevre and Picard [1994]; Moreno et al. [2004]; Nekovee et al. [2007]; Trpevski et al. [2010]).

Each of the above models may be used as a way of modelling the information diffusion in the social network with any kind of information transmitted. There are also rare cases when researchers were able to evaluate models against real world data, see Apolloni et al. [2009], Yang and Leskovec [2010], and Zbieg et al. [2012]. It shall be noticed that these models do not assume that an individual will be somehow personally attracted by the information - his or her personal state is irrelevant from the perspective of the whole process. This is the major difference when comparing the diffusion of information against the innovation diffusion or social influence, where the attitude or persuasion of an individual implies the state of neighbouring nodes. This and other differences will be described in Section 3.5.

3.2.3 The Role of Individuals in Information Diffusion

However, it should be remembered that apart from out-from-network sources, the information is transmitted (or not) by the decisions of individuals. They become familiar with information, rate the quality, importance and relevance of it and decide whether to inform the others or not. So another branch of research in this area is devoted to studying the information diffusion process at the lowest level - the motivations of people to become information spreaders or keepers. A very extensive work of Palloni [1998] summarises and justifies the sociological advances in this area, presenting also additional models of diffusion. Since the goal of this dissertation differs from presenting motivations of individuals in information diffusion, readers are advised to start with this survey if they are interested in extending their knowledge in this area.

3.3 Diffusion of Innovations

As it was already mentioned in Section 3.2, two concepts: diffusion of information and diffusion of innovations are distinguished. In the most cited book about diffusion of innovations, Everett M. Rogers states that *„Diffusion is the process by which an innovation is communicated through certain channels among the members of a social system. It is a special type of communication, in that the messages are concerned with new ideas"* (Rogers [2010]). This definition slightly contrasts with the point of view of the author, but for the purpose of this work it should be assumed, that the diffusion of information is represented as communication by Everett's definition and diffusion of innovations is the process of diffusion.

This definition leads to the conclusion that the diffusion of innovations among society results in social change of some type. To properly understand the idea of innovation, it should be remembered that the innovation itself must supersede some previous idea, i.e. it has some advantages of different kinds - economic, social, technological - that give it a chance to become accepted by an individual or a group of people. So, the innovation introduces a social change and the way this change occurs in an individual is presented in the next section. Here, in contrast to the previous section, which omitted the role of individuals in diffusion of information, it was decided to present the individual's perspective, since this aspect is more relevant to this dissertation.

3.3.1 The Innovation-Decision Process

Rogers [2010] presented a model of stages in the innovation-decision process. This process describes how a person becomes familiar with innovation, then how he or she makes the decision to adopt or reject it and, if adopted, how this person moves to the confirmation stage. It is an interesting study showing that this

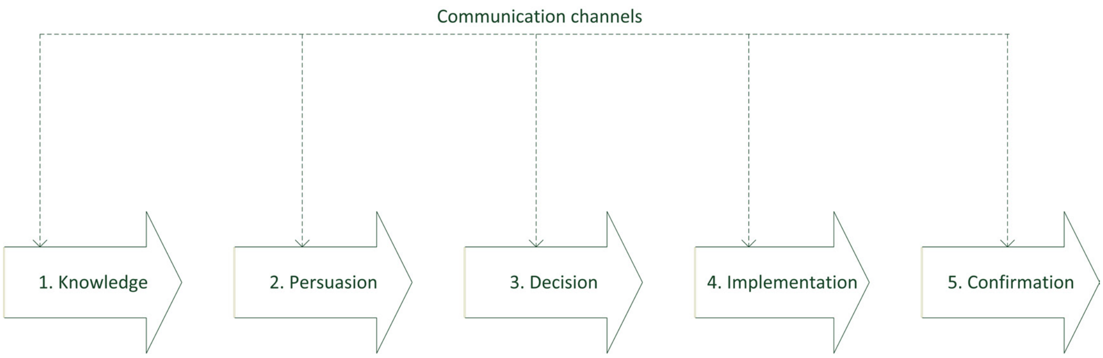

Figure 3.3: The model of innovation-decision process (Rogers [2010]).

process is complex and it involves many psychological and sociological aspects. The illustration of the model is presented in Figure 3.3 and each stage of it is described briefly below based on Rogers [2010].

1. Knowledge

The first stage, the knowledge stage, happens when an individual is exposed to the innovation's existence and gains some understanding of how it functions. It is often assumed, that this process is passive, that is a person does not actively look for a new innovation, but becomes informed about it by one of the communication channels (TV, peers, Internet). However, some researchers argue that this process requires activity, since the awareness of the innovation must be somehow initiated. An interesting part of this stage is also the question of how a person becomes aware of the need for this innovation. As the history of innovations shows, some innovations are the fulfilment of our needs, but some generate these needs. So the innovation may simply be better than its antecedent or, sometimes by being exposed to the innovation, we learn or discover new needs that have to be addressed and, indeed, these needs are addressed immediately by this innovation. The latter case is often described with the example of the Apple company. Their products often show to the customers some properties which instantly become highly desired. Here, the need for particular innovation is being born to become instantly fulfilled it by a particular Apple product (Carlson and Wilmot [2006]).

2. **Persuasion**

   After obtaining knowledge of the innovation, a person aligns it to his or her viewpoint, forming either a positive or a negative attitude towards it. Compared to the previous stage, persuasion is more related to the feeling rather than to the knowing, since attitude is a subjective concept and by having the same knowledge two people may form a different opinion of the innovation. Of course, since the knowledge of an innovation is rarely total, the uncertainty about it may lead some people towards forming a positive attitude anyway. They believe that it may fulfil their needs, and some others will try to avoid the risk and will present a more conservative approach - rejecting the innovation. Still, the main outcome of this stage is either a favourable or unfavourable attitude toward the innovation.

3. **Decision**

   As a person possesses the knowledge and the attitude towards the innovation, they either decide to adopt the innovation or to reject it in the decision stage. This part of the process is crucial from the social change point of view, since it is the moment when a person becomes convinced or not as to the innovation and performs some action. Rogers [2010] assumes that when a person decides to adopt the innovation, they also decide to make full use of it. On the other hand, rejection may be one of two kinds, *passive rejection* or *nonadoption*. The former means that the use of the innovation was never considered, and the *active rejection* (nonadoption) is the conscious decision not to adopt the innovation.

4. **Implementation**

   After making a positive decision about to adopt the innovation, an individual puts the innovation into use and this is the first moment when the social change becomes visible to others who may then be exposed to it. Until this part of the model, an individual was facing a mental exercise rather than real action about how he or she actively adopts the innovation.

5. **Confirmation**

   As Rogers [2010] states, some researchers finished the model on the *Im-*

plementation stage. However, it seems that innovation adopters often seek confirmation or reinforcement of their decision. They may even reverse their decision, if they become exposed to conflicting messages about the innovation. This is the stage where different kinds of dissonance may occur which should be eliminated or at least reduced. This dissonance comes from facing the attitudes of the closest neighbourhood towards the innovation adopted by an individual, from uncertainty about how to use the innovation or from other sources.

Last but not least, this model also involves *communication channels* by which different information is distributed. This information may be about the innovation itself, e.g. providing knowledge about it or about the adoption of the innovation by an individual or their peers. The communication channels are the same as the ones presented in Section 3.2, i.e. peers or out-of-network sources, such as mass media. These communication channels may be also considered as the sources of influence which affect individuals' decisions at every stage of the innovation-decision process.

3.3.2 Adopter Categories

The process of becoming an innovation adopter is time ordered, as it was shown in Section 3.3.1. Still, some stages may be shorter, the others longer, depending on the innovation type and the behaviour of an individual. Naturally, not all individuals adopt an innovation at the same time and this happens for various reasons. This differentiation led to the categorization of adopters which was also presented in Rogers [2010]. He enumerated five categories of adopters, referred to as *ideal types* and placed them on a Gaussian distribution plot, see Figure 3.4.

The understanding of these categories may be helpful in knowing the potential pathways of innovation across society, but it should be emphasised that this is an ideal categorization which may be not true for every innovation studied.

Innovators

The typical trait of innovators is venturesomeness. They are eager to learn about innovations, to try them, even if adopting the innovations will cost

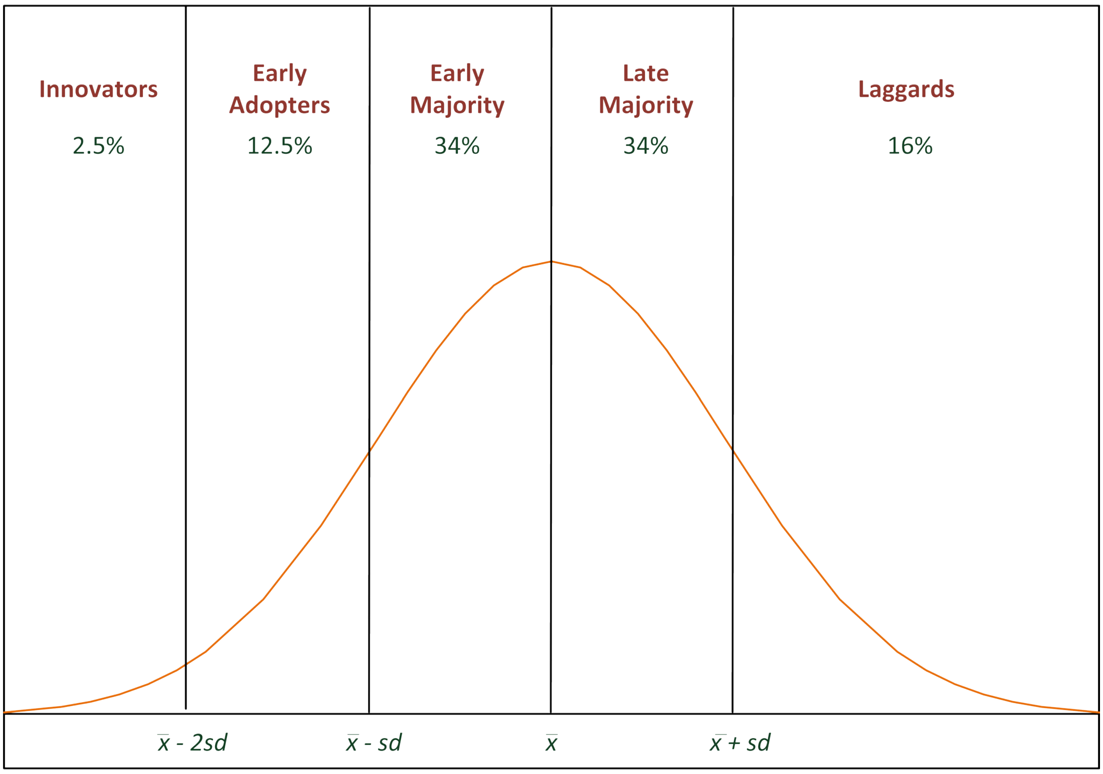

Figure 3.4: The *ideal types* representing adopter categories in the diffusion of innovations process (Rogers [2010]).

them more than an alternative, ans introduce risk. Despite the geographical distance, innovators form cliques that are communicating frequently. As Rogers [2010] states, they *play a gate-keeping role in the flow of new ideas into a social system.*

Early Adopters

Society needs people that are local to them to learn about innovations and this is the role of early adopters. They adapt the innovation and play the role of advisers in their environments with whom the others would like to discuss the potential of the innovation. Compared to innovators who may be described as cosmopolitans, early adopters are local leaders. They reduce the degree of uncertainty about innovations among their peers, playing the central role in their groups.

Early Majority

They adapt ideas just before the average member of the social system, but they do not hold a central position like early adopters, avoiding leadership.

Late Majority

The late majority are cautious about their decision and presents more often a sceptical opinion on the innovation. Members of this group more often require the motivation from others to be convinced as to the innovation, so the degree of uncertainty about the idea has to be reduced before they adopt it.

Laggards

Laggards represent cautiousness in adopting the innovation and this may be due to several factors. Their economic position may force them to avoid risky behaviour, before they can be sure that the innovation will become widespread. But some other reasons are also possible. For example, family traditions, or their position in the social network resulting in communication delay. It is often observed that when the laggard adopts one innovation, its successor is already adopted by innovators. There has been discussion as to whether the term *laggard* is not too pejorative when applied to this group, but no other has been agreed since the 1960s.

3.3.3 Models of Diffusion of Innovations

This subsection aims to present the three most popular mathematical models used to study the diffusion of innovations. They are presented in chronological order according to their introduction. An extended presentation of these and some other models can be found in Carrington et al. [2005], which also provides an extraordinary set of references to empirical studies on which these models were evaluated. These models may be divided into three categories:

- Macro models - assuming the perfect social mixing and no proximities between people.
- Spatial Autocorrelation - which reflect the physical distance.
- Network Models - focusing on the neighbourhood of an individual.

The first two models: the logistic and Bass model belong to macro models, whereas *Moran's I* belongs to spatial autocorrelation, and the last group covers the network models.

Parametric logistic model

This model, belonging to the group of macro models, is relatively simple, since it uses a one-parameter logistic function (Carrington et al. [2005]):

$$y_t = b_0 + \frac{1}{1 + e^{-b_1 t}}, \tag{3.1}$$

where y_t is the proportion of adopters to the whole population in time t, b_0 denotes the y intercept and b_1 is the parameter to be estimated. The simplicity of this model limited its applications, so it was substituted by the Bass model.

Bass Model

The Bass model is a two-parameter model that overcomes the obvious limitations of the one presented in Equation 3.1. It can be used to forecast expected reach of innovation or to estimate the diffusion rate. Its mathematical formulation is as follows (Bass [1969]):

$$y_t = b_0 + (b_1 - b_0)Y_{t-1} - b_1(Y_{t-1})^2, \tag{3.2}$$

where y_t is the proportion of adopters to the population in time t, b_0 is a rate parameter for innovation, Y_{t-1} is the number of adopters prior to the time t and a new parameter b_1 represents the rate parameter for imitation, i.e. the degree of adoption due to prior adopters. The most important improvement is that this model allows the incorporation of the percentage of adopters at each time point. The Bass model is widely used in empirical studies (e.g. Bass et al. [1994]; Dodds [1973]; Tigert and Farivar [1981]; Wright et al. [1997]) as well as often being extended to cover particular phenomena (Mahajan and Muller [1996]; Tseng and Hu [2009]).

Moran's I

Moran's I is an early model which tests spatial association, geographic clustering of adoption and it is based on the fact that it is relatively simple to obtain proximity data. The model is formulated as follows (Moran [1950]):

$$I = \frac{N \sum_i^N \sum_j^N d_{ij}(y_i - \overline{y})(y_j - \overline{y})}{S \sum_i^N (y_i - \overline{y})^2}, \tag{3.3}$$

where I indicates the level of autocorrelation, i.e. the level of clustering, N is the sample size, D represents the distance matrix and $d_{ij} \in D$, adoption is indicated by y, its mean value is $\overline{y}$, and S is the sum of all distances in the distance matrix. This approach measures the degree to which nodes that are connected to one another deviate from the average behaviour in the network similarly or differently (see Moran [1950] and the description in Carrington et al. [2005]). The main drawback of this approach is the lack of usage of network properties, i.e. it ignores the underlying network structure behind the diffusion process. That makes the Moran's I in some sense similar to the macro models, since none of them consider the actual network. The last approach tries to overcome this limitation by being strictly dependent on the network.

Network Models

The network-based approach uses a different point of view. Instead of looking from the global perspective at the diffusion process among society or the group, it analyses the ego (actor) networks (Prell [2011]; Valente [1996a,b]) to calculate the probability of adopting the innovation. It calculates the network exposure of an individual as the fraction of neighbours that adopted the innovation (Carrington et al. [2005]):

$$E_i = \frac{\sum w_{ij} y}{\sum w_i}, \tag{3.4}$$

where w is the weight of the network edge from node i to j and y is the vector of adoptions. This approach is the baseline for all the network-dependent models which incorporate the network structure for studying diffusion. Still, by being a very low-level one, it requires to know the topology of the network and obtaining this topology may be a challenging task.

Three different approaches for modelling the diffusion of innovations were presented in this section. They differ by point of view: the macro level attempts to estimate the outcome of the process without focusing on local properties of individuals, *Moran's I* uses the proximities and the network models represent the bottom-up approach starting from the node's perspective to end up with the global proportion of adopters.

These models may be applied to different situations, since it is unlikely that the researcher will have full knowledge of the environment where the diffusion process will take place. So when attempting to predict the outcome of the diffusion of innovations process, which model may fit the best must be known, depending on the knowledge of the innovation type and the social structure that adopts this innovation.

3.3.4 Summary

The above introduction to the process of diffusion of innovations plays an important role in this chapter. It shows how character, location in the social network and other aspects of an individual place him or her in the diffusion process. Moreover, the complexity of the individual decisions leading either to adoption or rejection an innovation is also presented. Since this was just an introduction to the far richer area combining psychology and sociology, this chapter now turns in the direction of the essence of the whole dissertation - the social influence which is a theory studying why people become influenced. It is definitely broader than the diffusion of information and presents a different approach than diffusion of innovations.

Concluding the topic of diffusion of innovations and just to demonstrate to the reader that it may be even more complex, some other competing approaches are mentioned. They were actively studied in Jensen [1983], Dunn and Gallego [2010] and Venkatraman et al. [1994].

3.4 Social Influence

3.4.1 Introduction

Social influence is defined as change in a person's cognition, attitude, or behaviour, which has its origin in another person or group (Raven [1964]). By using this definition it is relatively easy to find the basic difference with the diffusion of innovations. Here the major role in an individual's change is played by their neighbourhood whereas in the diffusion of innovations the main cause of social change was the innovation or product itself. Still, in most cases it is hard to distinguish what were the most important factors when adopting some innovation or attitude - peers, out-of-network sources or personal reasons. But when trying to compare these two concepts, the essential assumption is that in social influence the neighbourhood plays a more important role in adopting some behaviour than the subject, while diffusion of innovations puts the subject at the front of the decision process.

In this section the concept of social influence from the sociological perspective is presented. It is compared against diffusion of information and innovations to give the reader a full understanding of the rationale behind the differentiation of these concepts. It also introduces the models of influence; one of them will then be used in the research part of this thesis.

3.4.2 Sociological Background

Social influence is one of the major research areas in social psychology. It is a scientific study devoted to the observation of how people change each other's behaviour and attitudes. Another good definition of social influence apart from the one quoted in Section 3.4.1 appeared in Aronson [2003]: *„Social influence is the effect that people have upon the beliefs or behaviours of others"*. It is obvious that being a social species we are exposed to external influence, which ideally spreads among groups. On the other hand, some other phenomena, especially *homophily*, may be also important explanation why groups are somewhat similar (McPherson et al. [2001]). The influence may be of different kind: intentional (e.g. rewards or punishments) or unintentional (being a person considered important in a society). However the question is how people react to the social influence of others (Turner [1991]). Aronson distinguished three kinds of responses to social influence which better define why we tend to conform when exposed to the influence. These are: *compliance*, *identification* and *internalization*; they are briefly described below (Bagozzi and Lee [2002]; Kelman [1958, 1961]). These three types of reasoning behind adopting someone's point of view try to cover all the possible motivations of a person becoming influenced.

Compliance

This is the behaviour that makes human beings no different to other animals, since by behaving compliantly they are trying to gain some reward or avoid punishment. It is most often observed that this behaviour lasts as long as the promise of the reward or the threat. It is very unlikely to become a habit if these external motivations disappear (Cialdini and Goldstein [2004]) and a person has just the slightest or even no conviction at all that the idea of the influencer is right from this person's point of view.

Identification

Identification also represents a motivation in which it is unlikely that the influenced person is fully convinced of the idea of influencer. Instead, he or she is trying to become like the influencer, believing that this is the desired behaviour. Still, this phenomenon is likely to have a long-lasting effect, which differentiates it from *compliance*. So, the person starting to behave or think similarly to the influencer is most likely not intrinsically satisfied. The true reason is that it may lead to the positive relationship with the person or group (Aronson [2003]; Mugny et al. [1984]).

Internalization

Last but not least *internalization* is the most permanent response to social influence. Here the influenced person believes that the opinion, behaviour or attitude of an influencer is intrinsically right, so he or she accepts it internally, extending or modifying their own system of values. This is a true identification with the influencer, arising from the individual's commitment to the subject rather than from trying to adapt to someone else behaviour, gain rewards or avoid punishment. Moreover, due to this strong commitment, this person may even advocate this value to others, which is unlikely in the previously presented reactions (Lepper [1983]; McCauley [1989]; Ryan and Stiller [1991]).

These three responses to social influence are most likely met in human behaviour. However, sometimes it may happen that there is no single response type found, but adopting some value from others may be due to a combination of the above. Nevertheless, most likely, there is one crucial reason why someone is successfully influenced, the others being only side effects of it.

From the influencer perspective, it may be important to strengthen the effect of influence and different researchers try to provide methods of how this may be done. For instance, three factors increasing the likelihood of becoming influenced that are the part of *social impact theory* are provided in Latane [1981]:

- *Strength* - the importance of the influencer to the individual.
- *Immediacy* - physical and temporal distance of the influencer to the individual.
- *Number* - the size of the group influencing an individual.

Another work attempting to study the phenomenon of increasing the influence effect (Cialdini [2001]) recognizes the following techniques:

- *Reciprocity*, i.e. willingness to return a favour.
- *Commitment and Consistency* - being committed to the system of values.
- *Social Proof* - the need of being committed to the change by observing others behaviour.
- *Authority* - obeying authorities.
- *Liking* - willingness to follow the argumentation of people an individual likes.
- *Scarcity* - limiting resources may force people to act differently than they would without limited resources.

Of course, the above introduction to sociological background of social influence may be considered as limited, but it is not the main goal of this dissertation to provide all the reasons why people become influenced and how to maximize the influence in terms of sociology.

Now this chapter turns to presenting the most commonly used mathematical concepts for modelling social influence. Their goal is to model the process and at the end, to estimate the overall number influenced by making some prior assumptions taken from the sociological background.

3.4.3 Models of Social Influence

3.4.3.1 Introduction

Before presenting the most popular models of social influence in social networks, one sentence from Hedström and Bearman [2009] by D.J. Watts and P. Dodds may be mentioned: *„(...) it is still the case that formal models of social influence suffer from a dearth of realistic psychological assumptions"*. The problems of fitting real world data to models and trying to answer the question whether a particular influence process may be modelled with a selected approach still challenges researchers. The problem lies in the complexity of human behaviour and the impossibility of separating social processes that are occurring simultaneously. Still, many results achieved in this area tend to contradict this pessimistic point of view of Watts and Dodds. The continuous development of models or models' variations suggests that models will fit the reality much better in next few years (e.g. Aral and Walker [2012]). On the other hand, there still remains the gap between models and psychology that requires to be intensively studied to find the psychological rationale of particular behaviour expressed in the model.

The presented models of social influence may look similar to the network models presented in Section 3.3.3, since they focus on a network structure to answer the question whether someone will become influenced or not. This similarity is not accidental, since the point of view is identical. However, due to the fact that becoming influenced does not always mean adopting an innovation, different psychological processes may play an important role here. Moreover, social influence requires an external individual (an influencer). Because of those two reasons, in this thesis these two terms are distinguished. Further arguments are presented in Section 3.5.

Since the strength of social influence depends on many factors such as the strength of relationships between people in the networks, the network distance between users, temporal effects, characteristics of networks and individuals in the network (Sun and Tang [2011]), it is relatively hard to model all these factors combined. However, some research shows that under some assumptions there exist models that fit the reality well (Marsden and Friedkin [1993]; Masuda and Holme [2013]; Robins et al. [2001]). The main models that are most commonly used in this area are as follows:

- *The Linear Threshold model* (LT),
- *The Independent Cascade model* (IC),
- *The Voter Model* (VM),
- *The Naming Game* (NG).

Each of them tries to consider the psychological background, but as it was previously stated, sometimes it is just a loose interpretation of humans' behaviour, that still fits the reality. For these models their recent variants which are suitable for some real world cases are also presented.

From the historical perspective, studying the social influence in terms of analytical process was the case of trying to model how influence spreads in time. Starting from a set of influenced nodes in time t_0 which are in this work denoted as $\Phi(0)$, as time unfolds, more and more neighbours of $\Phi(0)$ become influenced if they fulfil the model criteria. Most typically, these processes are modelled in directed graphs and focus on a *progressive* case, where nodes may become *influenced* from an *uninfluenced* state, not the other way round (Kempe et al. [2003]). Since this is a network approach, the influence process runs through edges in the graph and most typically no other external factors of influence are considered, such as out-of-network sources.

3.4.3.2 The Linear Threshold Model

The most recognizable model for social influence is Granovetter's Linear Threshold Model (Granovetter [1978]), but a similar approach was also proposed by Schelling [1978]. In this model, a node v is influenced by each v's neighbour w according to a weight $b_{v,w}$, such that $\sum_{w \in N_v^{inf}} b_{v,w} \leqslant 1$. Then each node v chooses a *threshold* θ_v from the interval $[0, 1]$ and this threshold represents the value which has to be overcome by the aggregation of v neighbours' influence in order to influence the node v. So the formal condition of influencing the node v is as follows:

$$inf(v) = \begin{cases} 1, & \text{when} \quad \sum_{w \in N_v^{inf}} b_{v,w} \geqslant \theta_v \\ 0, & \text{when} \quad \sum_{w \in N_v^{inf}} b_{v,w} < \theta_v, \end{cases} \tag{3.5}$$

where $inf(v)$ is the result of influence process for node v, θ_v denotes the threshold level for node v, N_v^{inf} the set of influenced neighbours of node v and $b_{v,w}$ represents the influence weight of node w on node v. The influence process ends when no more nodes can be influenced - this is the stop condition.

In this case, the value of θ_v represents the individual's chances of becoming influenced when enough of its neighbours are influenced. So all the psychological factors are included in this parameter and it should be also underlined that this approach represents the individual's perspective rather than the influencer perspective. Granovetter illustrated the model with the hypothetical case of a riot.

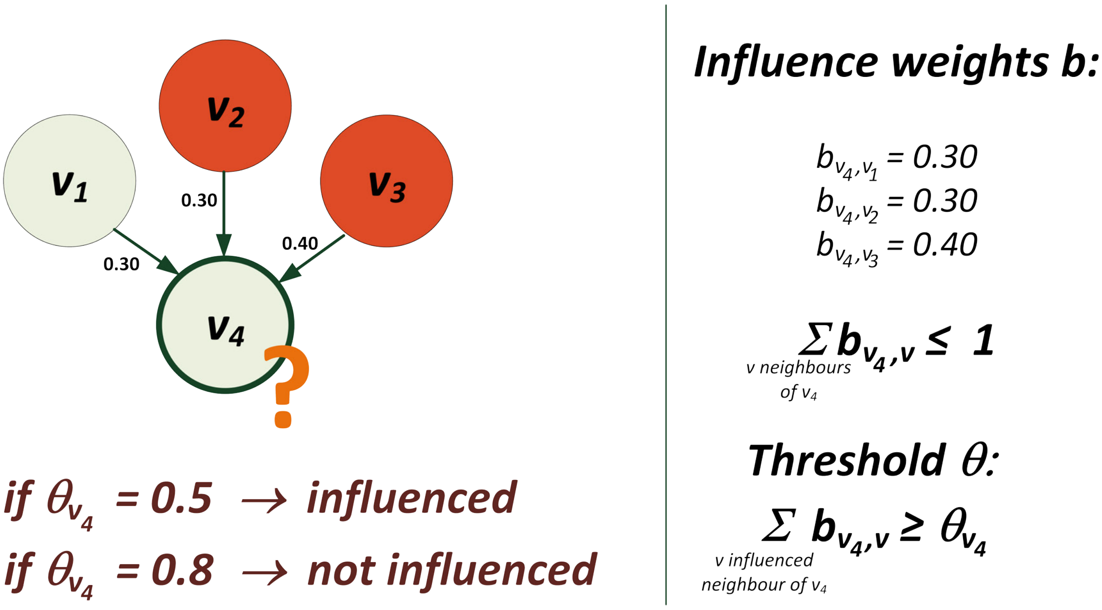

Figure 3.5: An illustration of the Linear Threshold model.

Since individuals were unsure what the costs and benefits of joining it are, they observed their peers and considered joining only when enough of their neighbours joined the riot. Otherwise, they refrained.

Of course, the greatest question is how to assign particular values of θ to individual nodes. There are two most typical approaches: draw them from a probability distribution $f(\theta)$ (Granovetter [1978]) or hard-wiring them at a fixed value (Berger [2001]; Peleg [1997]). The most interesting and realistic scenario is the former, i.e. drawing θ_v from a distribution, since the distribution represents both the average tendencies and also the heterogeneity present in the population. Lowering or raising the mean of $f(\theta)$ would modify the general susceptibility of the population, while increasing or decreasing the variance would correspond to an increase or decrease in variability in susceptibility among individuals (Hedström and Bearman [2009]). Still, hard-wired thresholds are also often considered in science. An illustration showing how the LT model works is presented in Figure 3.5 and the model is formalized as Algorithm 1 (based on Zafarani et al. [2014]).

The LT model became a core of many modifications or extensions. For instance, Goyal et al. [2010] extended this model by introducing temporal decay, as well as factors such as the influence-ability of a specific user, and influence-

Algorithm 1 Linear Threshold model

Require: Graph $G(V, E)$, set of initially influenced nodes $\Phi(t_0)$, thresholds θ_v, influence weights $b_{v,w}$

return Final set of influenced nodes $\Phi(K)$
k = 0;
Uniformly assign random thresholds θ_v from the interval [0, 1];
while $k = 0$ or $\Phi(t_{k-1}) \neq \Phi(t_k)$ **do**
    $\Phi(t_{k+1}) = \Phi(t_k)$;
    $uninfluenced = V \setminus \Phi(t_k)$;
    for all $v \in uninfluenced$ **do**
        if $\sum_{w \text{ influenced neighbour of } v} b_{v,w} \geqslant \theta_v$ **then**
            influence v;
            $\Phi(t_{k+1}) = \Phi(t_{k+1}) \cup \{v\}$;
        end if
    end for
    $k = k + 1$;
end while
$\Phi(K) = \Phi(k)$;
Return $\Phi(K)$;

proneness of a certain action. On the other hand, Barbieri et al. [2013] proposed topic-aware extensions of the LT model. In Pathak et al. [2010] authors considered multiple cascades of the LT model and they allow nodes to switch between them, whereas Borodin et al. [2010] introduced a number of modifications to the competing model variant: they force nodes to draw one cascade they join at the end of the process or consider the mutual influence of cascades on each other.

3.4.3.3 The Independent Cascade Model

The next model has its roots in interacting particle systems (Durrett et al. [1988]; Liggett [1985]) and it is called the Independent Cascade model - IC (Goldenberg et al. [2001]; Kempe et al. [2003]; Król [2014]). Again, the process starts with a set of influenced nodes $\Phi(0)$, but each node v in the network has assigned a probability $p_{v,w}$. According to this probability the node v has a single chance to influence its neighbour w and if it fails, it will have no other chance. If it succeeds, w will become influenced in the next step. The process runs until no more influences are possible. The IC model is presented as Algorithm 2 (based on Zafarani et al. [2014]).

Algorithm 2 Independent Cascade model

Require: Graph $G(V, E)$, set of initially influenced nodes $\Phi(t_0)$, activation probabilities $p_{v,w}$

return Final set of influenced nodes $\Phi(K)$
k = 0;
while $\Phi(t_k) \neq \{\}$ **do**
  $k = k + 1$;
  $\Phi(t_k) = \{\}$
  for all $v \in \Phi(t_{k-1})$ **do**
    for all w neighbour of v, $w \notin \cup_{j=0}^{k} \Phi(t_j)$ **do**
      $rand$ = generate a random number in [0,1];
      if $rand < p_{v,w}$ **then**
        influence w;
        $\Phi(t_k) = \Phi(t_k) \cup \{w\}$;
      end if
    end for
  end for
end while
$\Phi(K) = \cup_{j=0}^{k} \Phi(t_j)$;
Return $\Phi(K)$;

In the case of this model, from a psychological perspective, the influencer becomes more important, since he or she holds the probability $p_{v,w}$ and this is one of the differences between IC and LT models. In the LT model, the influence process parameter was assigned to the uninfluenced node and in the IC model it is held by the influencer. Just as in the previous model, the probability may be fixed or drawn from a distribution $f(p)$.

Again, there are many variants of the IC model. The already mentioned work by Barbieri et al. [2013] introduced the topic-aware approach also for this model, while Kempe et al. [2005] studied the *decreasing cascade model*. One of the problems with the base LT and IC models is that they need to assume the influence probabilities and there are works that try to obtain these probabilities from past propagations. It may not be considered as an extension of the base model, but an approach to make the probabilities or threshold more realistic. One of the works in this area is Saito et al. [2008]. There also exists an approach to model multiple independent cascades in the network (Bharathi et al. [2007]).

3.4.3.4 The Voter Model and the Naming Game

An interesting case of influence in networks is the case where two separate opinions or influences are competing in society. This phenomenon may be observed in many situations and it has its roots in studying the consensus processes (Lu et al. [2009]) or the language dynamics (DallAsta et al. [2006]). Below there are two variants of the process presented: the Voter Model (VM) and the Naming Game (NG) .

The Voter Model introduced in Clifford and Sudbury [1973] and extensively analysed later in Holley and Liggett [1975] assumed that each node in the network can hold one of two opinions and by interacting with others it may switch the opinion to the opinion of the peer. The model also introduced the degree of conformity which defines whether a node will follow the majority (conformist) or minority (non-conformist), see Javarone [2014].

On the other hand, the Naming Game, also referred to as the binary-agreement model (Xie et al. [2011]) introduced another variant of forming the opinion or spreading the influence. At any time a node may possess one of two competing opinions or two opinions simultaneously. In a given step in time, we choose a node randomly, designate it as the speaker, choose one of its neighbours randomly and designate it as the listener. The speaker proceeds to convey its opinion to the listener (selected randomly if it possesses two). If the listener possesses this opinion already, both speaker and listener retain it while eliminating all other opinions; otherwise, the listener adds the opinion to his list (Xie et al. [2012]).

Both of these models are useful in studying common phenomena occurring in social networks which involves binary options, such as reaching the consensus on contradictory opinions or observing which of the competing parties will win the election. The current research trends suggest that these models will be actively studied and extended in the future (e.g. Jankowski et al. [2012a]; Li et al. [2013]; Maity et al. [2013]; Mobilia [2013]; Rogers and Gross [2013]; Zhang et al. [2013]).

3.4.3.5 Models of Influence - Summary

The above presented models are just a selection of models which allow the study of the influence process in networks analytically. As it was presented, they differ by the perspective (LT versus IC), and by the number of competing influences (VS and NG versus others), but all of them are linked to the same process - social influence. Sometimes their applicability is limited, but the empirical research shown, so far demonstrated that they model human behaviour accurately in some cases (Marsden and Friedkin [1993]; Robins et al. [2001]), even if the psychological background of an individual is more complex than just a single parameter.

At this point it is worth comparing the above presented phenomena dealing with the diffusion or social influence in social networks. It may help the reader to see the similarities and differences between them. This comparison is presented in Section 3.5. It will be followed by the global perspective - the spread of influence in social networks.

3.5 Comparison of Diffusion and Influence Processes

Table 3.1 lists all of the discussed social phenomena presented so far: diffusion of information, diffusion of innovations and social influence. They are briefly summarized to provide a quick overview of similarities and differences between them.

While summarizing these three approaches or theories, there are two questions that should be addressed. Firstly, why are there so many similarities between diffusion and innovations and social influence? In this dissertation diffusion of innovations is rather a process of becoming convinced and committed by the individual itself. This person is observing how his family or relatives react to a given innovation, whether they adopt it or not, what opinion about it is presented in mass media? This is why individual factors play the major role here. On the other hand, the process of social influence is mostly invoked by others and the individual eventually becomes influenced, or not. So diffusion of innovations is rather an internal process becoming convinced of some idea or innovation, while social influence introduces some pressure towards a person that is about to become influenced.

Secondly, why do researchers use the term *diffusion* for two processes: diffusion of innovations and diffusion of innovations? If using the physical definition (Philibert [2005]), diffusion is *the movement of a substance particles* and there is no possibility that the number of particles will increase. So, when looking at this definition it is rather hard to justify the usage of this term, since while sharing some information we do not loose it. On the other hand, the process of diffusion assumes that there is *movement from the region of high concentration to a region of low concentration* and it might be key to proper understanding of the reasons for this name. Here, highly concentrated information or ideas (just a number of seeds or sources) diffuse towards regions with low concentration, reducing local maxima of density.

Table 3.1: Comparison of diffusion and influence processes in social networks.

	Diffusion of information	**Diffusion of innovations**	**Social influence**
Short description	General process of transmitting the information between individuals or between out-of-network sources and individuals. It includes any kind of information. The role of out-of-network sources is important.	This process reflects the diffusion of innovation or idea through the network.	Social influence is defined as change in a person's cognition, attitude, or behaviour, which has its origin in another person or group.
The content of transmission	Any information, regardless of the content.	Idea, attitude, innovation.	Since the process of social influence does not focus on the product or innovation, the content may be anything that can a man be influenced with.
Models	Birth-death processes, branching processes, epidemic models.	Parametric logistic model, Moran's I, Bass model, network models.	Linear threshold, independent cascades, naming game, voter model.
Application	Speeding up or controlling the information flow over the network, especially social media application.	Marketing products or promoting desired behaviour.	Politics, marketing, social behaviour, daily habits.

3.6 Spread of Influence in the Social Network

In Section 3.4, a number of properties of the social influence process were presented. Overall, social influence is the process where an individual becomes influenced by others to some idea, behaviour or attitude and the role of external entities is underlined here. However, as this process starts with a single influencer or a group of them, we can see that despite the fact that the influence itself is rather an individual case, it spreads among others at a network level. It means that it may be observed as a complex psychological and sociological process when thinking about the reasons why an individual is becoming influenced, but when taking another perspective the whole outcome of the influence process may be studied. This outcome is referred to as the *spread of influence*, i.e. the reach of the influence in a given network. It is most often measured by the number of influenced nodes after the process ends according to the fixed number of initially influenced individuals. The spread of influence for social network and temporal social network is defined formally below.

Definition 5. *The Spread of Influence* $SI^{SN}(SN, m, P^m, SC, \Phi(0))$ *for a social network* SN*, a given propagation model* m *and its parameters* P^m*, stop condition* SC *and initial seed set* $\Phi(0)$ *is the total number of influenced nodes from* SN *after the stop condition* SC *is reached:* $SI^{SN}(SN, m, P^m, SC, \Phi(0)) = |\Phi|$*. Here,* Φ *denotes the set of influenced nodes. Since the spread of influence* $SI^{SN}(SN, m, P^m, SC, \Phi(0))$ *is considered for a fixed propagation model* m*, its parameters* P^m *and stop condition* SC*, it will be further denoted as* $SI^{SN}(SN, \Phi(0))$*. To shorten the further textual content,* $SI^{SN}(SN, \Phi(0))$ *will be referred to as SI.*

Spread of influence for social networks is an iterative process, where for each iteration i it is possible that new nodes become influenced. The stop condition SC in Definition 5 refers to the definition of a given propagation model m. As it is presented in Section 3.4.3.2, the LT model reaches its stop condition after the iteration for which no more nodes can be influenced. The other propagation models may use different stop condition definitions, e.g. a fixed number of iterations.

Definition 6. *The Spread of Influence* $SI^{TSN^K}(TSN^K, m, P^m, p, \Phi(0))$ *for a temporal social network* TSN^K*, a given propagation model* m *and its parameters*

P^m, period T_p and initial seed set $\Phi(0)$ is the total number of influenced nodes $SI^{TSN^K}(TSN^K, m, P^m, p, \Phi(0)) = |\Phi(T_p)|$, where $\Phi(T_p)$ denotes the set of influenced nodes at the end of period T_p. Since the spread of influence $SI^{TSN^K}(TSN^K, m, P^m, p, \Phi(0))$ is considered for a fixed propagation model m, e.g. LT, it will be further denoted as $SI^{TSN^K}(TSN^K, p, \Phi(0))$.

In Figure 3.6, the reader can observe the process of the spread of influence when a given social influence model was selected, so both terms are related to each other, but in fact they refer to different perspectives. The distinction between these two terms is often ignored, but it is important for this dissertation. In Chapter 4, more information about spread of influence can be found, since this dissertation focuses on maximizing it.

3.7 Summary

The goal of this chapter was to present three important social phenomena and theories: diffusion of information, diffusion of innovations and social influence. Despite that they may look similar, each of them refers to different aspects of information, innovation or influence transmission over society. The main difference between them is the perspective or point of view. Diffusion of information is the most general approach and it does not consider the individual factors as strongly as latter two theories. Indeed, models for studying the diffusion of information ignore the personal factors or at least minimize them to some extent. In contrast, the diffusion of innovations and social influence focus on individuals while looking for the answer to why particular ideas, opinions, innovations or attitudes spread over society. However, what really differentiates them is the way people become convinced or influenced. In the former, diffusion of innovations, is mostly explained as a combination of individual aspects of a person that commits to some idea. In the latter - social influence - it is mainly the role of the influencer that persuades a person towards something and this pressure may be subtle or not. The rest of this dissertation will focus on social influence.

In next chapter, an important problem of social influence will be discussed. The research question that is addressed there and has been studied for more than

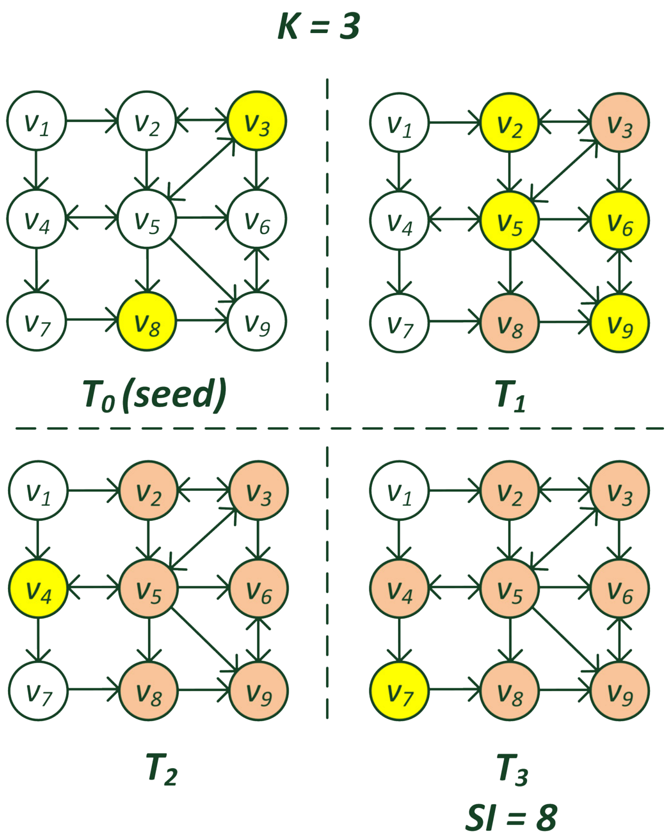

Figure 3.6: An exemplary social influence process following the linear threshold model in a social network SN. The threshold value is fixed for all nodes, the weights are equal, $\theta_v = 0.33$. At the beginning $\Phi(t_0) = \{v_3, v_8\}$, at the end of the process $\Phi(t_3) = V \setminus \{v_1\}$ where V denotes the set of vertices of SN, v_1 cannot be influenced for this combination of parameters for the LT model.

a decade is how to maximize the outcome of the influence process, the spread of influence. This chapter covers the definitions of the research problem for both cases: time-aggregated and temporal social networks, as well as the state-of-the-art in this area.